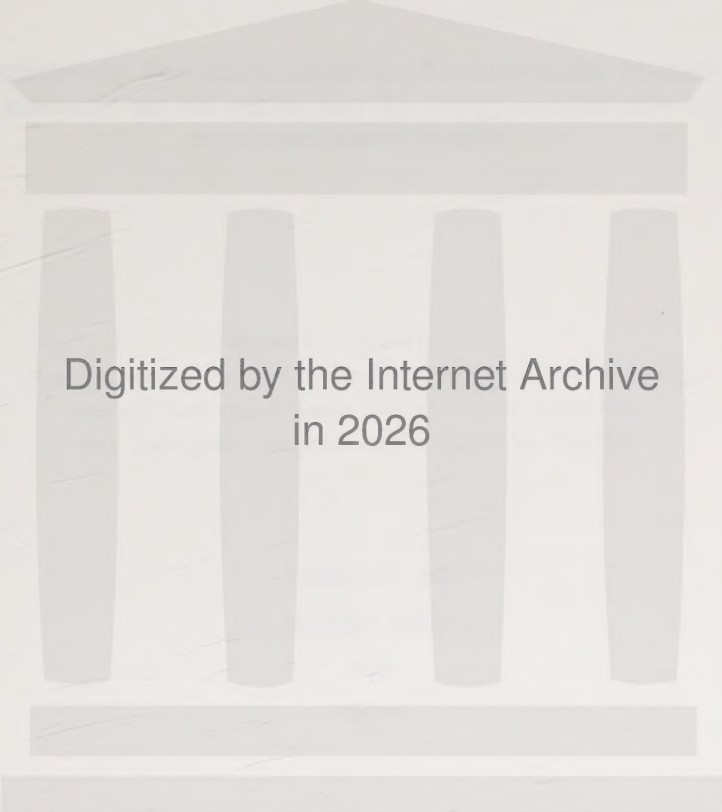


**UNTERSUCHUNG ÜBER DIE DIAGNOSTISCHE
ÜBEREINSTIMMUNG ZWISCHEN KLINIKERN UND
DEM SYSTEM DIAL ZUR
COMPUTERUNTERSTÜTZTEN
PSYCHIATRISCHEN DIAGNOSTIK**

Wilhelm Schmid

1974



Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41551921_0112

Aus dem Max-Planck-Institut für Psychiatrie

München

(Direktor: Professor Dr. Detlev Ploog)

UNTERSUCHUNG ÜBER DIE DIAGNOSTISCHE ÜBEREINSTIMMUNG ZWISCHEN
KLINIKERN UND DEM SYSTEM DIAL ZUR COMPUTERUNTERSTÜTZTEN
PSYCHIATRISCHEN DIAGNOSTIK

Inaugural - Dissertation

zur Erlangung der Doktorwürde in der gesamten Medizin
verfaßt und einer Hohen Medizinischen Fakultät
der Ludwig-Maximilians-Universität zu München

vorgelegt von

WILHELM SCHMID

aus

Prien

1973

München

1. Berichterstatter: Prof. Dr. v. Zerksen
2. Berichterstatter: Prof. Dr. Ploog
Dekan: Prof. Dr. Spann

Tag der mündlichen Prüfung: 23.12.1974



In Dankbarkeit
für meinen Lehrer Friedrich Lechner

Herrn Professor Dr. von Zerssen danke ich für die Überlassung des Themas und seine Ratschläge und Verbesserungen bei der Durchführung der Arbeit. Auch Herrn Dr. Kombour, Herrn Dr. Castell und Herrn Dr. Mittelsten Scheid bin ich sehr dankbar für ihre freundliche Unterstützung, ebenso Frau Koeller und Frl. Köhler, sowie Frau Irmann und den Locherinnen, die an der Durchführung der Computerdiagnostik beteiligt waren.

INHALT

	Seite
1. <u>Einführung</u>	5
1.1. Problemstellung	5
1.2. Literaturübersicht	8
1. Allgemeiner Teil	8
2. Spezieller Teil: DIAGNO, CAPPS, DIAGNO II, DIAL	12
2. <u>Methodik</u>	16
2.1. Auswahl der Fälle	16
2.2. Die Diagnostiker	17
2.3. Die Dokumentationssysteme: IMPS, AMP	17
2.4. Technischer Ablauf	19
2.5. Schwierigkeiten	
1. bei der Auswertung der Dokumentationssysteme	19
2. bei der Übertragung der Information in die CAPPS	20
3. bei der Diagnostik mit Hilfe von Symptomlisten	21
2.6. Statistische Auswertung	21
3. <u>Ergebnisse und Diskussion</u>	22
4. <u>Zusammenfassung</u>	25
5. <u>Literaturverzeichnis</u>	26

1. Einführung

Die moderne Computertechnik findet in der Medizin bereits 1.1. vielfach Anwendung. So macht man sich etwa in der Labordiagnostik, der Hämatologie oder der Kardiologie die Fähigkeit der Elektronenrechner zunutze, gewaltige Mengen an Information speichern und beliebig vergleichen zu können. Auch in der Psychiatrie wird seit etwa zehn Jahren, besonders in England und den Vereinigten Staaten, an der Entwicklung computerunterstützter Diagnostik gearbeitet. Der Fortschritt solcher Bemühungen liegt wohl weniger in der Schaffung von "Diagnosenmaschinen", die eines Tages den Arzt weitgehend ersetzen sollen, als vielmehr in der Möglichkeit, über die Standardisierung der Nomenklatur und des diagnostischen Prozesses einen internationalen Erfahrungsaustausch und epidemiologische Vergleichsstudien durchzuführen und die psychiatrische Ausbildung zu vereinheitlichen. In der vorliegenden Untersuchung, die am Max-Planck-Institut für Psychiatrie in München durchgeführt wurde, wird ein in den Vereinigten Staaten entwickeltes Programm für psychiatrische Computediagnostik auf seine Verwendungsfähigkeit hierzulande geprüft. Vorweg jedoch einige grundsätzliche Anmerkungen zur Diagnostik in der Psychiatrie.

In der Psychiatrie herrscht weltweite Uneinigkeit über die Einteilung der Krankheiten. "Kategorialen" (Kraepelin) und "dimensionalen" nosologischen Systemen (Katz, Eysenck), die den Kranken einer diagnostischen Kategorie zuordnen (nomothetischer Ansatz), steht die "ideographische" Auffassung einer individualisierenden Betrachtung des Patienten gegenüber, die auf diagnostische Einteilung weitgehend verzichtet (Psychoanalyse, "Behaviorismus").³³⁾

Innerhalb der nosologisch-klassifikatorischen Richtung gibt es wiederum verschiedene, voneinander abweichende Schulrichtungen, deren teils deskriptive, teils ätiologische Einteilungen uneinheitlich "motiviert" sind und kaum je Phänomenologie, Ätiologie und Prognose gleichmäßig berücksichtigen. Daher wurde schon vorgeschlagen, für jeden dieser Aspekte

ein eigenes Klassifikationsprinzip einzuführen (Zigler et al.)³⁴⁾ oder jede Diagnose obligat dreiteilig anzugeben:

1. nosologische Kategorie, 2. Hauptsyndrom, 3. wesentliche Ätiologie (Essen-Möller and Wohlfahrt)⁶⁾.

Die Kommunikationsschwierigkeiten und die mangelhafte Übereinstimmung in der psychiatrischen Diagnostik sind nach einer internationalen Studie von Ward et al.²⁹⁾ zu mehr als 60% auf Unterschiede in den nosologischen Schemata (vgl. Meyer¹⁶⁾; Mombour¹⁷⁾) zurückzuführen. Einen ersten Schritt zur Überwindung dieser "nosographischen Sprachverwirrung" (Gross)¹¹⁾ stellt die Einigung auf den psychiatrischen Diagnosenschlüssel der WHO (ICD = International Classification of Diseases) dar, dessen 8. Revision (dtsh. Ausgabe von Degkwitz et al.)⁷⁾ seit einigen Jahren in allen der WHO angeschlossen Ländern für offizielle Gesundheitsstatistiken gültig ist.

Auch die Diagnosen des Computerprogramms, das in der vorliegenden Arbeit getestet wurde, sind ICD-Diagnosen. Die Einteilung unterscheidet zunächst symptomatologisch zwischen

1. krankhaften Zuständen mit schwerer Beeinträchtigung der psychischen Funktionen, die zu Störungen des Realitätsbezugs führen (Psychosen und Demenzen),
 2. nicht-psychotischen psychischen Störungen ohne Beeinträchtigung der Intelligenz (Neurosen, Psychopathien, sexuelle Deviationen, Süchte, psychosomatische und andere psychogene Störungen, nicht-psychotische psychische Störungen auf organischer Grundlage),
 3. geistiger Entwicklungshemmung (Oligophrenie).
- Diese drei Gruppen werden dann nach Ätiologischen und symptomatologischen Gesichtspunkten weiter unterteilt.

Ein weiterer Schritt in Richtung auf das Ideal einer weltweiten Übereinstimmung in der Verwendung diagnostischer Begriffe ist deren Operationalisierung durch ein allgemein verbindliches Glossar. Ein solches wurde für den psychiatrischen Teil der ICD von einem Expertenteam der WHO erarbeitet (vgl. Degkwitz et al.)⁷⁾. Das Glossar der ICD enthält jedoch keine strikten Regeln zur Standardisierung des diagnostischen Vorgehens, so daß trotz seiner Verwendung die Diagnostikerübereinstimmung oft gering ist (vgl. Kraus)¹²⁾.

Zubin gibt einen Überblick über eine Reihe von Arbeiten, die sich mit der Reliabilität (Zuverlässigkeit; ausgedrückt in der Beobachterübereinstimmung) und Validität (Treffsicherheit, Genauigkeit der Klassifikation) psychiatrischer Diagnosen befassen.³⁵⁾ Die festgestellte Übereinstimmung ist unterschiedlich und teilweise sehr gering, z.B. für depressive Neurose 18% - 63%, für Psychopathie 6% - 66%. Daran sind nicht nur die in der Psychiatrie häufigen "atypischen" und "Grenzfälle" schuld, sondern nachgewiesenermaßen auch der Einfluß der Persönlichkeitsstruktur des Untersuchers und seiner Haltung dem Patienten gegenüber, vor allem jedoch die Anwendung unterschiedlicher Explorationstechniken und verschiedenartiger Strategien im diagnostischen Entscheidungsprozeß.³⁾⁴⁾⁶⁾⁹⁾²⁴⁾

Die Standardisierung der Informationsverarbeitung setzt aber eine standardisierte Informationsgewinnung voraus. Eine Vielzahl von genormten Untersuchungsinstrumenten wurde für diesen Zweck entwickelt (vgl. Mombour)¹⁸⁾. Im einfachsten Fall hat der Untersucher im Anschluß an die Exploration eines Patienten eine Symptomliste abzuheften. Differenzierter sind Schätzskalen, in denen die einzelnen "Items" (in der Regel Symptome) abgestuft nach ihrem Ausprägungsgrad mit einem Punktwert (score) versehen werden. Einen hohen Grad an Objektivität gewährleisten Fragebogen mit standardisiertem Interview, in denen Reihenfolge und Formulierung der an den Patienten zu stellenden Fragen vorgeschrieben sind.

Die in solcher Form standardisiert vorliegende Information über einen Patienten läßt sich codieren und in einen Computer eingeben. Es besteht jedoch eine grundsätzliche Schwierigkeit in der "Reinheit" und "Randunschärfe" dieser Daten: Im Gegensatz zu "harten", d.h. genau definierten Merkmalen wie Größe, Gewicht oder Temperatur, und noch relativ harten Daten wie EKG und EEG sind die aus einem psychiatrischen Interview gewonnenen Angaben über Vorgeschichte und Befund eines Patienten weniger präzise objektivierbar und daher für den Computer nicht gleich gut geeignet.¹¹⁾

Um aus diesen Informationen eine Diagnose abzuleiten, muß im Computer der diagnostische Entscheidungsprozeß eines Psychiaters nachvollzogen werden. Gauron et al.⁹⁾ untersuchten den recht undurchsichtigen intrapsychischen Vorgang bei der Diagnosenfindung und stellten sechs verschiedene Vorgehensweisen fest, die sich unterscheiden 1. in der Methode der Informationsgewinnung (strukturiert, starr - unstrukturiert, flexibel) und 2. im Denkprozeß des Psychiaters (induktiv, logisch - intuitiv, alogisch). Zwei dieser sechs Wege werden allgemein in der Computerdiagnostik beschritten und seien deshalb hier erwähnt. Der eine kombiniert ein starres Schema der Informationsgewinnung mit logischem Denkprozeß: Erst nach möglichst vollständiger Sammlung aller einschlägigen Informationen wird ein Vergleich der gefundenen Merkmalskonstellation mit typischen Krankheitsbildern vorgenommen und dasjenige herausgesucht, mit dem ein möglichst hoher Grad an Übereinstimmung besteht (Mustererkennung oder "pattern recognition"). Der andere Weg ist flexibler: Anfangs werden mehrere Diagnosen in Betracht gezogen, die dann durch systematisches Abfragen der jeweils benötigten Information der Reihe nach ausgeschlossen werden, bis die passende Diagnose "per exclusionem" gefunden ist.⁹⁾³³⁾ Die von Spitzer und Endicott entwickelten Computerprogramme DIAGNO²⁷⁾ und DIAGNO II²⁸⁾ gehen diesen Weg.

Die Regeln für solche standardisierte Informationsverarbeitung (Algorithmen; siehe Lange¹⁴⁾) lassen sich beim Programmieren als einzelne logische und rechnerische Schritte (Operationen) codieren und in sog. Flußdiagrammen ("flow charts") schematisch darstellen. Hierzu wurden eigene problemorientierte Sprachen entwickelt, z.B. die in Spitzers DIAGNO-Programmen verwendete FORTRAN ("Formular Translation"). Ehe ich jedoch auf DIAGNO und DIAL genauer eingehe, sei zunächst auf die Arbeiten einiger anderer Autoren hingewiesen.

Overall und Gorham verfaßten einen standardisierten Untersuchungsbogen, die "Brief Psychiatric Rating Scale" (BPRS), bestehend aus 16 Symptom-Einheiten wie Angst, Schuldgefühle, Mißtrauen usw., die in einem siebenstufigen Ausprägungsgrad

angegeben werden. Eine Reliabilitätsprüfung der einzelnen Items ergab für verschiedene Untersucherpaare eine Übereinstimmung von 0,56 bis 0,87.²⁰⁾ Dieser Korrelationskoeffizient K (Kappa), der nachfolgend noch mehrfach erwähnt wird, ist die beobachtete Übereinstimmung, korrigiert um die zu erwartende zufällige Übereinstimmung.

Für geringere als die zufallswahrscheinliche Übereinstimmung wird K negativ (bei -1 Maximum an Diskordanz); bei rein zufälliger Übereinstimmung, d.h. ohne systematischen Zusammenhang zwischen den beiden untersuchten Variablen, ist K gleich Null; für zunehmende Konkordanz kann K positive Werte bis maximal +1 annehmen.

$$K = \frac{p_o - p_c}{1 - p_c}, \text{ wobei } p_c = \text{zu erwartende zufällige Übereinstimmung}$$
$$p_o = \text{beobachtete Übereinstimmung}$$

(vgl. Spitzer et al.)²⁶⁾

Anhand der 16 Items aus der EPRS stellten 19 Psychiater unabhängig voneinander "typische" Symptomenmuster auf für 13 verschiedene Diagnosen aus dem APA-Manual^{*)}). Die Übereinstimmung dieser "Symptom-Cluster" war sehr hoch, auch in einer anschließend mit weiteren neun Psychiatern und zehn klinischen Psychologen durchgeführten Wiederholung des Versuchs.²⁰⁾

Der Vergleich zwischen den so gewonnenen Standardbefundmustern und dem Befundmuster eines Patienten läßt sich auf einer elektronischen Rechenanlage durchführen. Allerdings ist bei der Aufstellung der Standardbefundmuster der Schluß von einer Krankheit auf eine bestimmte Symptomkombination ebensowenig eindeutig wie der umgekehrte Vorgang bei der Diagnostik, da eine Krankheit mehrere Befundmuster aufweisen kann und umgekehrt. Durch Gewichtungungsverfahren, die die Sensibilität (Häufigkeit, mit der das Symptom bei einer bestimmten Krankheit auftritt) und Spezifität (Fehlen des Symptoms bei nicht oder anders Erkrankten) berücksichtigen, erhält jede Diagnose einen bestimmten Grad an Wahrscheinlichkeit bei einem vorgegebenen Symptommuster.¹⁴⁾

*) APA = American Psychiatric Association¹⁾

Programmiert man einen Computer mit diesem "Conditional Probability Model", so wird ein "echter" Patient, der ja in vielen Fällen gerade nicht dem typischen Bild einer Krankheit entspricht, vom Computer derjenigen diagnostischen Kategorie zugeordnet, deren "klassischen" Zeichen er am meisten ähnelt.

Denselben Weg der Mustererkennung wählte Smith in seinem Programm.²⁵⁾ Er erweiterte Overalls Modell von 16 auf 41 Symptomeinheiten und von 13 auf 38 diagnostische Kategorien. Von 14 Klinikern wurden jeder der 38 Diagnosen alle 41 Symptome in vierstufigem Ausprägungsgrad zugeordnet. Computerdiagnosen, die auf diesem Wahrscheinlichkeitsmuster beruhen, stimmten in der Unterscheidung großer diagnostischer Kategorien wie Psychose, Neurose oder Psychopathie zu 87% mit Klinikern überein, bei einer allerdings recht kleinen Stichprobe (30 Fälle).

Nathan schlägt ein diagnostisches Modell vor, das bereits in der Form linearer "Flußdiagramme" konzipiert ist und sich deshalb für die elektronische Datenverarbeitung besonders anbietet.¹⁹⁾ Eine entsprechende Anwendung des Modells ist allerdings bislang nicht erfolgt.

Der diagnostische Prozeß gliedert sich in folgende sechs Abschnitte: 1. Unterscheidung Psychose/Neurose, 2. abnormes psychomotorisches Verhalten (quantitativ und qualitativ), 3. abnorme Wahrnehmungen (Illusionen, Halluzinationen), 4. abnormes Denken (nach Form, Ablauf und Inhalt der Gedanken; Gedächtnisstörungen), 5. affektives Verhalten (gehobene Stimmung, Depression, Spannung, abnormer Affektausdruck), 6. Bewußtseinsstörungen.

Wing, der mit seinen Mitarbeitern eine zunächst einfache deskriptive Einteilung der chronischen Schizophrenie vorgenommen hatte, entwickelte daraus unter Hinzunahme von Neurosen und akuten endogenen Psychosen²⁶⁾ eine Symptomliste von ca. 600 Items, das "Symptom Inventory" oder "Schedule for Psychiatric Present State Examination"²⁰⁾, dessen 9. Revision inzwischen auch ins Deutsche übersetzt worden ist.^{*)} Alle Angaben beziehen sich auf den Zeitpunkt des Interviews und den vorangehenden Monat (Anamnesezeiten werden

*) M. v. Cranach, Psychiatrische Universitätsklinik München

in einer getrennten Untersuchung erhoben). Für die einzelnen Fragen an den Patienten wird meist eine bestimmte Formulierung vorgeschlagen (standardisiertes Interview). Die Items sind zu 17 Gruppen zusammengefaßt (Angst, Depression, Gedächtnisstörung etc.), von denen jede zunächst durch "Testfragen" angeschnitten wird. Wenn kein Hinweis für eine Abnormalität in der betreffenden Richtung besteht, geht der Untersucher an einem sog. "cut-off"-Punkt zur nächsten Gruppe über. Auf diese Weise wird in einem durchschnittlichen Interview immer nur eine Teilmenge aller Fragen gestellt.

Die 600 Items beschreiben 145 klinische Symptome, die nach der Anzahl und dem Ausprägungsgrad zutreffender Items gewichtet werden. Die Symptome sind zu 35 "klinischen Syndromen" zusammengefaßt, die zu einer diagnostischen Zuordnung des Patienten führen. Diese "provisorische" Kategorisierung ist noch keine endgültige Diagnose, da keinerlei Information aus der Vorgeschichte verwendet wird.

In einer Reliabilitätsprüfung untersuchten fünf Psychiater paarweise 172 Patienten und diagnostizierten eine der 11 möglichen provisorischen Kategorien. Dabei erreichten sie in 33,7% völlige Übereinstimmung, in 7% teilweise Übereinstimmung (Erstdiagnose identisch, Zweitdiagnose verschieden) und in 9,3% Nichtübereinstimmung. Die Anwendungsmöglichkeiten ihres Instruments sehen die Autoren vor allem in Forschungsarbeit und Ausbildung; ein Programm zur Computordiagnostik (CATEGO) befindet sich noch im Stadium der Erprobung. Die Übereinstimmung zwischen Kliniker- und CATEGO-Diagnosen lag bei 56,1% (7. Revision) und 75,7% (8. Revision). Dabei handelte es sich allerdings fast ausschließlich um Fälle von Schizophrenie, und die klinische Diagnose stellte ein bzw. seine Mitarbeiter.³¹⁾

Cassano berichtet sehr gute Erfolge mit seinem Programm P.D.A. (Procedimento Diagnostico Automatico) an der Universität Pisa.⁵⁾ Dabei werden die LIPS-Fragebogen von Lorr et al.¹⁵⁾ benutzt (s.w.u.) und für jeden Patienten mit faktoranalytischen Methoden ein Symptom-Profil erstellt. Durch Vergleich mit gespeicherten Profilen gelangt der Computer

zu einer oder mehreren Diagnosen unter Angabe des Sicherheitsgrades in Prozent. Die Werte für jeden Patienten bleiben gespeichert, wodurch das Material, auf das sich die Diagnostik stützt, ständig anwächst. Eine Untersuchung mit 514 Patienten zeigte Übereinstimmung mit Klinikerdia gnosen in 76%.

Am Biometric Research Institute in New York entwickelten Spitzer und Endicott im Laufe der letzten Jahre eine Reihe von standardisierten Untersuchungsinstrumenten und die Programme DIAGNO und DIAGNO II zur Computerdiagnostik. 1.2.2.

Bei DIAGNO²⁷⁾ werden dem Computer die Informationen über den Patienten mittels des "Psychiatric Status Schedule" (PSS) geliefert, das aus einer Liste von 492 Items und einer standardisierten Interviewanleitung besteht. Das Programm faßt alle Items zu klinisch geläufigen Syndromgruppierungen zusammen. Ein "Entscheidungsbaum" (logical decision tree) mit 36 Verzweigungen führt durch differentialdiagnostische Erwägungen "per exclusionem" zur Diagnose. Dabei wird der Reihe nach erst auf organische Ursache, dann auf Psychose, auf Psychopathie und schließlich auf Neurose geprüft. Das Ergebnis jeder Frage bestimmt die nächstfolgende und scheidet dabei u.U. eine oder mehrere Diagnosen oder Diagnosengruppen aus. Innerhalb jeder der großen Gruppen wird dann weiter differenziert, gegebenenfalls auch zu einer vorher bereits ausgeschlossenen Gruppe zurückgesprungen.

Der Computer stellt auf diese Weise eine von 25 Standarddiagnosen der APA¹⁾, zuzüglich der Diagnosen "nicht krank" und "unspezifische Krankheit mit leichter Symptomatik". Beim Vergleich mit den Diagnosen von 30 Klinikern ergab sich, daß DIAGNO "theoretisch konstruierte" Fälle ebenso gut erkannte wie die Ärzte; aber bei 200 "echten" Fällen, die die Kliniker ebenfalls ausschließlich nach PSS-Protokollen zu beurteilen hatten, stimmten die DIAGNO-Diagnosen nicht so gut mit den Psychiatern überein wie diese untereinander.²⁷⁾

Die PSS-Protokolle bezogen sich nur auf die Symptomatik des Patienten zum Zeitpunkt der Exploration und in der vorangehenden Woche. Um durch die Hinzunahme zusätzlicher Information aus der Lebensgeschichte und Krankheitsanamnese

die Übereinstimmung zu verbessern, entwickelten Spitzer und Endicott die "Current and Past Psychopathology Scales" (CAPPS). Dabei kombinierten sie eine erweiterte Version der "Psychiatric Evaluation Form" (PEF) zur Erfassung des gegenwärtigen Zustandes eines Patienten mit dem "Psychiatric History Schedule" (PHS), das zur Dokumentation der Vorgeschichte dient (vgl. Mombour¹⁸). Der Fragebogen besteht aus 130 Items in meist sechsfacher Abstufung über die Vorgeschichte vom 12. Lebensjahr an (Anamnese) und 41 Angaben über den Befund zum Zeitpunkt der Exploration und im vorangehenden Monat (Befund).

Eine Faktorenanalyse, die im wesentlichen der Datenreduktion durch Syndrombildung gleichkommt, ergab 16 Faktoren für den A-Teil (Anamnese) und 7 Faktoren für den B-Teil (Befund). Der A-Teil erfaßt: Depression/Angst, antisoziale Haltung, intellektuelle Leistung, Realitätsbezug, Wahnhaftigkeit/Zwanghaftigkeit, organischer Schaden, Manie, gestörte soziale/sexuelle Beziehungen, Körperbezogenheit, Hysterie, sexuelle Perversion, Sinnesstörung, Abhängigkeit, Widerspenstigkeit/Argwohn, Absonderlichkeit, Gereiztheit. Der B-Teil erfaßt: psychotische Störung, antisoziale Einstellung, Depression/Angst, Abkapselung, Körperbezogenheit, Sinnesstörung, Wahnhaftigkeit/Phobie.

An 31 Patienten wurde eine Reliabilitätsprüfung durchgeführt. Je zwei Beobachter, die zusammen einer Exploration beigewohnt hatten, füllten anschließend unabhängig voneinander die Fragebogen aus und erzielten eine sehr hohe Übereinstimmung (K um 0,90 und höher).

Von den 171 Items der CAPPS werden 94 Items (plus Alter und Geschlecht) im DIAGNO II - Programm verwendet. Der Computer erstellt daraus eine von 44 APA-Diagnosen¹⁾, außerdem die Diagnosen "nicht krank" und "unspezifische Krankheit mit leichter Symptomatik". Die Diagnosen der APA entsprechen weitgehend den ICD-Diagnosen; der Hauptunterschied liegt darin, daß im Diagnosenschema der APA organische Psychosen in akute und chronische Formen aufgeteilt werden.

Auf dem Manual der APA¹⁾ beruht auch die Entscheidungslogik von DIAGNO II: Zuerst werden Zeichen abgefragt, die auf eine organisch bedingte Hirnfunktionsstörung hindeuten (Alkohol- und Drogenabusus, Cerebralsklerose). Falls dafür kein Anhalt vorliegt, wird sukzessive weitergesucht nach

Hinweisen für endogene Psychosen, für vorübergehende situative Reaktionen, Störungen des Sozialverhaltens, Persönlichkeitsstörungen, Neurosen, psychosomatische Erkrankungen, geistigen Entwicklungsrückstand und schließlich für unspezifische Erkrankung. Jede einmal in Betracht gezogene Diagnose kann an späteren Entscheidungspunkten wieder verworfen werden. Wie bei DIAGNO (I) wird innerhalb der Kategorien weiter differenziert. Das Programm wurde an 52 echten und 92 konstruierten Fällen getestet und laufend modifiziert, bis weitere Veränderungen keinen Effekt mehr zeigten. Das endgültige Programm hat 57 Entscheidungspunkte und ist in FORTRAN IV für die IBM 360 geschrieben.

Zur Messung der diagnostischen Übereinstimmung zwischen Psychiatern und Computer wurde das sog. "weighted kappa" (K_w) (Sandifer)^{*)} in leicht modifizierter Form verwendet. Sandifer hatte ein siebenstufiges Maß für die "Ähnlichkeit" bzw. "Verschiedenheit" psychiatrischer Diagnosen eingeführt (0 = exakte Übereinstimmung bis 6 = totale Verschiedenheit). Beim Vergleich einer Serie von Diagnosenpaaren läßt sich dann die Gesamtübereinstimmung in K_w ausdrücken:

$$K_w = 1 - \frac{\sum w_i P_{oi}}{\sum w_i P_{ci}}, \text{ wobei } w_i = \begin{array}{l} \text{Nichtübereinstimmungsgrad} \\ \text{eines Diagnosenpaares} \end{array}$$

P_{oi} = beobachtete Übereinstimmung
 P_{ci} = zufallsbedingte Übereinstimmung

Für zwei Diagnostiker, die in allen Fällen identische Diagnosen stellen, ist $K_w = 1$.

22 Psychiater erstellten Interviewprotokolle von 100 Patienten auf CAPPS und konstruierten außerdem für jede der 40 möglichen Computerdiagnosen ein "hypothetisches" Protokoll. Die Daten wurden auf modifizierte CAPPS-Sogen übertragen, in denen jegliche Information ausgespart wurde, die von DIAGNO II nicht benutzt wird. Acht Klinikern wurde dann paarweise je ein Viertel der Protokolle zur Diagnostik vorgelegt. Die Übereinstimmung war diesmal - im Gegensatz zu DIAGNO (I) - auch für die echten Fälle zwischen Computer und Klinikern ebenso gut wie zwischen den Klinikern untereinander. Es ergab sich

^{*)} 1966, unveröffentlicht; zit. nach Spitzer and Endicott²⁶⁾.

ein durchschnittliches K_w von 0,45. Die Kappas der Interviewer, die die CAPPS ausgefüllt hatten, waren jeweils etwas niedriger, d.h. sie stimmten mit den übrigen Klinikern und mit dem Computer weniger gut überein. Das läßt sich nach Spitzer²⁸⁾ erklären aus der zusätzlichen Information, die den Interviewern zur Verfügung stand. Außerdem diagnostizierte DIAGNO II häufiger als die Kliniker "nicht psychisch krank" und "unspezifische Krankheit". In diesen Fällen hatten die Kliniker Symptome berücksichtigt, die hinter den "cut-off levels" des Computerprogramms lagen.

Zusammenfassende Übersicht:

	DIAGNO	DIAGNO II
"input":	39 Items	94 Items
berücksichtigter Zeitraum:	Woche der Exploration	ab 12. Lebensj.
Entscheidungspunkte:	36	57
Diagnosen:	27	46
Übereinstimmung mit Klinikern:		
konstruierte Fälle:	$K_w = 0,60$	$K_w = 0,61$
echte Fälle:	$K_w = 0,28$	$K_w = 0,45$

Hier stimmte zum ersten Mal der Computer mit Klinikern ebenso gut überein wie diese untereinander, wenn allen dieselbe Information zur Verfügung stand. Zum ersten Mal bestand auch beträchtliche Übereinstimmung mit den Diagnosen von Klinikern, die den betreffenden Patienten genau kannten.

Die hohe Übereinstimmung, die Smith²⁵⁾ berichtet (87%), ist nach Ansicht Spitzers schwierig zu interpretieren, da Smith, der selbst an der Entwicklung des Computerprogramms beteiligt war, sowohl den Fragebogen ausfüllte, als auch die klinische Diagnose stellte.

In der vorliegenden Arbeit soll nun geprüft werden, inwieweit das Programm DIAGNO II auf unsere diagnostischen Vorstellungen übertragbar wäre und ob es zu Ausbildungszwecken und Fallkonferenzen verwendet werden könnte. Dazu ist es notwendig, die diagnostische Übereinstimmung des

Computers mit Klinikern, die den jeweiligen Patienten selbst exploriert hatten oder denen Symptomlisten über den betreffenden Fall vorgelegt wurden, zunächst "off-line" zu überprüfen. Ein "on-line"-Programm (DIAL) in Form eines Frage-Antwort-Dialogs zwischen Kliniker und Computer, bei dem der Entscheidungsprozeß vom Kliniker verfolgt und korrigiert werden kann, wurde am Max-Planck-Institut für Psychiatrie, München, von Gerster und Dirlich¹⁰⁾ aus dem DIAGNO II - Programm entwickelt. Das DIAL-Programm ist so allgemein gefaßt, daß es auf beliebige Skalen und beliebige Entscheidungsbäume anwendbar ist. In der vorliegenden Untersuchung wird es im "off-line"-Verfahren getestet, wobei die CAPPS verwendet werden und der geringfügig modifizierte Entscheidungsprozeß des DIAGNO II - Programms zugrunde liegt.

Die Diagnosen des Programms wurden dem ICD-Schema angepaßt und ihre Anzahl auf 50 erhöht. Außerdem wird im Knoten 120 für die Entscheidung, ob weiter in Richtung auf endogene Psychose oder aber auf Neurose hin abgefragt wird, der Schwellenwert angehoben. In diesem Knoten wird nach dem Vorhandensein von Wahnideen und Halluzinationen gefragt. War nur eines dieser beiden Symptome "angedeutet vorhanden" (Ausprägungsgrad 2), so wurde - vor unserer Änderung - bereits eine endogene Psychose diagnostiziert. Da auch bei neurotischen Erkrankungen paranoide Symptome "angedeutet vorhanden" sein können, haben wir den Schwellenwert für die Summe der Werte beider Symptome von 2 auf 4 erhöht.

2. Methodik

Wir verglichen die Diagnosen, die an denselben Fällen von verschiedenen "Diagnostikern" gestellt wurden. Bei den Fällen handelte es sich um 98 Patienten (zwei von ursprünglich 100 Krankengeschichten wurden aus technischen Gründen ausgeschieden), die 1969/70 auf der psychiatrischen Abteilung des MPI stationär behandelt worden waren. Die Fälle waren nach folgenden Gesichtspunkten ausgewählt worden: Erstens sollte die vorläufige Diagnose, die im Anschluß an das Aufnahmeinterview gestellt worden war, mit der Entlassungsdiagnose übereinstimmen. Unklare Fälle oder Patienten, bei denen während des stationären Aufenthalts die Diagnose revidiert werden mußte,

2.1.

wurden ausgeschieden. Zweitens sollten möglichst alle der im DIAL-Programm vorgesehenen Diagnosen vertreten sein.

Die "Diagnostiker", deren Beurteilungen miteinander verglichen wurden, waren

2.2.

- a) diejenigen Kliniker, die den betreffenden Patienten während seines Klinikaufenthalts untersucht und die Diagnose in der Krankengeschichte (KG) niedergelegt haben;
- b) zwei "Fragebogendiagnostiker", d.h. zwei Kliniker, die nur mit der Information aus den CAPPS, also ohne persönliche Kenntnis der Patienten, auskommen mußten (FD1 und FD2);

FD1 ist Facharzt für Psychiatrie mit langjähriger klinischer Erfahrung und seit 1966 an der Abteilung für Erwachsenenpsychiatrie am Max-Planck-Institut für Psychiatrie tätig. FD2 - ebenfalls Facharzt für Psychiatrie - war zur Zeit der Untersuchung vier Jahre am Max-Planck-Institut für Psychiatrie, davon ein Jahr an der Abteilung für Erwachsenenpsychiatrie, tätig gewesen. FD1 und FD2 stellten anhand der CAPPS-Listen unabhängig voneinander ihre Diagnosen.

- c) ein Computer, dem die gleiche standardisierte Information über die Patienten zur Verfügung stand und der die diagnostische Entscheidung mit Hilfe des DIAL-Programms durchführte. Er wählte aus den 50 in dem Programm vorgegebenen ICD-Diagnosen jeweils eine Haupt- und ggf. eine Zusatzdiagnose aus.

Die Kliniker des MPI hatten nach der Patientenexploration jeweils zwei standardisierte Fragebogen zur Dokumentation des psychopathologischen Befundes ausgefüllt: die IMPS (Inpatient Multidimensional Psychiatric Scale) von Lorr et al.¹⁵⁾ und Blatt 3 ("Psychischer Befund") des AMP-Systems²²⁾.

2.3.

Die IMPS stellt eine Liste mit 90 Items dar, die sich auf die Symptomatik des Patienten zum Zeitpunkt der Exploration und in der vorangehenden Woche beziehen, also weiter zurückliegende anamnестische Angaben außer Acht lassen. Die Antworten werden entweder als "ja/nein" oder in fünf- bzw. neunstufigem Ausprägungsgrad angegeben.

Mit faktorenanalytischen Methoden faßten die Autoren diese Items zu zehn Syndromen zusammen: Erregung, feindselige Einstellung, paranoide Projektion, Größenwahn,

Wahrnehmungsstörungen, Ängstliche Selbstbestrafungstendenzen, Hemmung und Apathie, Desorientiertheit, motorische Störungen, Denkstörungen.

Die Übereinstimmung zwischen verschiedenen Untersuchern in der Beurteilung der einzelnen Items lag zwischen $K = 0,61$ und $K = 0,92$ (durchschnittlich $0,77$), bei Addition zu Summenscores für die oben angeführten Faktoren erreichte sie sogar Werte zwischen $0,82$ und $0,91$.¹⁷⁾ Nach diesen Faktorenscores lassen sich Profile zeichnen, die bestimmten Krankheitstypen zugeordnet werden können.

Das zweite verwendete Dokumentationssystem ist der Logen Nr. 3 ("Psychischer Befund") der "Arbeitsgemeinschaft für Methodik und Dokumentation in der Psychiatrie" (AMP), zu der sich 1965 elf deutsche, schweizerische und österreichische Universitätskliniken zusammengeschlossen haben. Die Antwortlisten sind so abgefaßt, daß sie direkt in ein Ablesegerät eingegeben werden können.

123 psychopathologische Items sind zu Gruppen zusammengefaßt: Bewußtseinsstörungen, Orientierungsstörungen, Aufmerksamkeits- und Gedächtnisstörungen, formale Denkstörungen, überwertige Ideen/Zwänge/Phobien, Wahn, Sinnestäuschungen, Ich- und Persönlichkeitsstörungen, Verstimmungen, Gefühlsstörungen, psychomotorische Störungen, Verlaufsbesonderheiten, Störungen des Trieb- und Sozialverhaltens/Pflegebedürftigkeit, Intelligenz.

In einem Manual²²⁾ werden die Items genau definiert; die Zeitperiode, auf die sich die Angaben beziehen sollen, ist allerdings nicht klar begrenzt. Jedes einzelne Symptom wird entweder als "nicht beurteilbar", "fraglich vorhanden", "nicht vorhanden", "leicht", "mittel" oder "schwer" ausgeprägt eingestuft. Genaue Reliabilitäts- und Validitätsprüfungen sowie faktorenanalytische Untersuchungen werden z.Zt. an der Psychiatrischen Abteilung des Max-Planck-Instituts für Psychiatrie vorgenommen.

Die Information für den B-Teil (Befund) der JAPPS konnte zu einem gewissen Grad aus diesen beiden Dokumentationssystemen entnommen werden. Die Angaben für den A-Teil (Anamnese) entstammen der Krankengeschichte des jeweiligen Patienten. Die Krankengeschichten waren von 21 verschiedenen Klinikern verfaßt und folgendermaßen aufgebaut: Familienanamnese, äußere Lebensgeschichte, frühere Erkrankungen,

innere Lebensgeschichte, Vorgeschichte der jetzigen Erkrankung, psychischer Befund, Diagnose, Differentialdiagnose, weitere diagnostische Maßnahmen und Behandlungsplan.

Der technische Ablauf war folgendermaßen: Nach Auswahl 2.4.
geeigneter Fälle füllte ich für jeden Patienten einen CAPPS-Antwortbogen aus nach den Angaben in der jeweiligen Krankengeschichte und den Eintragungen auf den IMPS und den AMP-Bogen. Von den CAPPS wurden die Daten auf Lochkarten übertragen und dem mit DIAL programmierten Computer zur Diagnostik eingegeben. Dieser lieferte außer einer Haupt- und ggf. einer Zusatzdiagnose auch eine Zusammenstellung der CAPPS-Items eines jeden Patienten, geordnet nach dem Ausprägungsgrad. Anhand dieser Listen stellten die beiden "Fragebogen-diagnostiker" unabhängig voneinander ihre Diagnosen.

Das Codieren der Krankengeschichtsinformationen war mit 2.5.1.
folgenden Schwierigkeiten verbunden: Die standardisierten Befunderhebungsbogen AMP und IMPS waren von beschränktem Wert, da sie keine Anamnesedaten enthalten und auch die zum gegenwärtigen Befund in den CAPPS geforderten Items nicht alle aufgeführt sind (z.B. die Fragen 27: antisoziale Haltung, 42: Krankheitsschweregrad, 43: Krankheitsdauer, 44: Auslösefaktoren, Stress). Zur Beantwortung dieser Fragen wurde ebenfalls die Krankengeschichte herangezogen. Weiterhin werden Symptome, die zum Zeitpunkt der Exploration verschwunden sind, in der Regel nicht mehr aufgeführt. So wird beispielsweise im AMP-Bogen ein vorangegangener Suizidversuch, der zur Einweisung in die Klinik geführt hatte, nicht erwähnt, wenn bei der Exploration keine Suizidtendenzen mehr erkennbar sind. Außerdem können beim AMP-Bogen ganze Symptomgruppen en bloc als "nicht zutreffend" abgehakt werden, wobei nachträglich nicht feststellbar ist, ob die Items tatsächlich einzeln geprüft wurden oder ob eines nicht beachtet wurde. Gelegentlich bestanden erhebliche Abweichungen zwischen den Angaben der IMPS und der AMP-Bogen zum selben Patienten. In diesen Fällen berücksichtigte ich in der Regel diejenige Angabe, die das betreffende pathologische Item als vorhanden bzw. in stärkerer Ausprägung dokumentierte. Die Beobachtungsperiode für den B-Teil erweiterte ich von vier auf zwölf

Wochen vor dem Zeitpunkt der Exploration, um die "jetzige Krankheitsperiode", die häufig weiter als vier Wochen zurückging, von der Vorgeschichte abzusetzen.

Der Anamneseteil der CAPPS enthält eine Reihe von Items, 2.5.2. die nicht regelmäßig in den Krankengeschichten erwähnt werden (etwa 45: Freundschaften in der Adoleszenz, 76: Freundschaften im Erwachsenenalter) oder nicht sicher beurteilt werden konnten. So ist 95: Anzahl früherer Krankheitsperioden, oder 96: Anzahl früherer depressiver und manischer Phasen schwierig anzugeben, wenn ein chronischer Verlauf oder psychopathologische Störungen zwischen den Krankheitsperioden vorliegen.

Die CAPPS-Items sind nicht immer sinnvoll definiert, z.B. 76: Freundschaften im Erwachsenenalter: 1 = ausgezeichnet, ... hatte zumindest einen sehr engen Freund, dem er vertraute...; 4 = mäßig, ... hatte einige spezielle Freunde, vermied aber Gruppenunternehmungen...; dazwischen liegt 2 = sehr gut und 3 = recht gut. Ein anderes Beispiel sind die Fragen 47 und 48: Sexualverhalten in der Adoleszenz und als Erwachsener, wo nach der angebotenen Definition etwa ein Jugendlicher, der bis zum Alter von 18 Jahren "kaum körperliche Beziehungen zum anderen Geschlecht" hatte, bei "homosexuellen oder sonst sexuell abartigen Verhaltensweisen" eingestuft wird. (Die Definition dieser beiden Fragen wurde von uns geändert.) So gibt es eine ganze Reihe von Items, in die zu einem gewissen Grad die Sittlichkeitsauffassungen und die Weltanschauung des Untersuchers mit eingehen, wie 56: Kriminalität ("Berücksichtigen Sie Art und Häufigkeit der Vergehen und Straftaten!"), 57: antisoziales Verhalten, oder 82: mangelndes Verantwortungsgefühl. Wo immer solche subjektiv gefärbten Feststellungen bei der Diagnostik berücksichtigt werden, ist mit verminderter diagnostischer Übereinstimmung zu rechnen.³⁾

Schließlich ist es auch anhand einer ausführlichen, detaillierten Krankengeschichte oft schwierig, Auslösefaktoren (98), Remissionsgrad (97) und Schweregrad (99) früherer Krankheitsepisoden zuverlässig festzustellen.

Einige der geschilderten Schwierigkeiten werden sich bei künftiger klinischer Anwendung des Systems verringern, wenn anstelle von Krankengeschichte und Befundbogen die direkte Exploration des Patienten die Grundlage bildet.

Für die beiden "Fragebogendiagnostiker" bedeutete die Beschränkung auf die CAPPS den Wegfall der "nichtverbalisierten Information", die der Kliniker beim persönlichen Kontakt mit dem Patienten gewinnt. Wie ungern sich Psychiater auf die schriftlich festgelegten Beobachtungen und Interpretationen von Kollegen verlassen, beschreibt Gauron⁹⁾ in seiner Untersuchung des diagnostischen Vorgehens.

2.5.3.

Die für die Diagnostik von Psychopathien notwendigen fremdanamnestischen Angaben, die der Kliniker aus dem Gespräch mit Angehörigen und Bekannten des Patienten erhält, werden in den CAPPS nicht ausreichend festgehalten. Die Listen stellen ein sehr "trockenes" Material dar, und die Diagnostiker fanden es nicht immer leicht, gleichzeitig viele, oft konträre Angaben vor Augen, sich ein plastisches Bild vom Patienten zu machen. Dies würde erleichtert durch eine Zusammenfassung von Items, z.B. alles, was sich auf die emotionale Struktur des Patienten bezieht oder auf die Kindheit bzw. die Zeit zwischen Pubertät und Krankheitsausbruch. Die Angaben über die Anzahl früherer Krankheitsperioden (95) und den Krankheitsverlauf (100) sind nicht nur, wie oben erwähnt, unsicher zu erheben, sondern auch schwierig zu bewerten.

Zur statistischen Auswertung wurden die 50 möglichen DIAL-Diagnosen zu Gruppen zusammengefaßt, die in drei aufeinanderfolgenden Ansätzen immer weiter aufgefächert wurden. In der ersten Einteilung wurde zunächst grob unterschieden zwischen Psychose und Nicht-Psychose (Neurose, Psychopathie u.s.w.). Psychotische und neurotische Depression wurden dabei eingeklammert und mit "reaktiver depressiver Psychose" (ICD-Nr. 298.0) zusammengefaßt zu einer dritten Gruppe "Depression". Die restlichen Kategorien waren Oligophrenie, "nicht psychisch krank" und "unklarer Fall".

2.6.

In der zweiten Einteilung wurden die Psychosen in körperlich begründbare und endogene aufgeteilt und die Süchte von

den Neurosen und den übrigen nicht-psychotischen Störungen abgetrennt. Die anderen Gruppen blieben gleich.

Die dritte Einteilung differenzierte genauer und unterschied die folgenden Kategorien: Alkoholpsychosen, sonstige Psychosen bei organischen Erkrankungen, Schizophrenie, schizoaffective Psychose, Manie, Depression, Neurosen ohne Depression, Persönlichkeitsstörungen, sexuelle Verhaltensabweichungen, Alkoholismus, sonstige Süchte, psychosomatische Erkrankungen, nicht-psychotische psychische Störungen bei körperlichen Erkrankungen, Oligophrenie, nicht psychisch krank, unklarer Fall.

Von den vier Diagnosen, die zu jedem Patienten vorlagen, wurde zunächst jede mit jeder anderen verglichen und ihre Übereinstimmung innerhalb der drei Einteilungen festgestellt. Sodann untersuchte ich, wie sich der Übereinstimmungsgrad bei den einzelnen diagnostischen Kategorien unterschied. Danach wurden neben den Erstdiagnosen auch noch die Zweitdiagnosen berücksichtigt.

Die Übereinstimmung wurde jeweils in Prozentwerten ausgedrückt. Die Verwendung eines speziellen Übereinstimmungskoeffizienten, z.B. K (vgl. S. 9) oder K_w (vgl. S. 14), schien in dieser Untersuchung keinen Vorteil zu bringen, da es nur auf einen relativen Vergleich der paarweisen Übereinstimmung ankam. Bei Verwendung von K_w ("weighted kappa") wird bei "ähnlichen" Diagnosen die Teilübereinstimmung berücksichtigt, was im allgemeinen eine zahlenmäßig höhere Übereinstimmung ergibt als das von uns verwendete Verfahren.

3. Ergebnisse und Diskussion

Die Übereinstimmung der Erstdiagnosen von Krankengeschichte, DIAL, FD1 und FD2 ist in Tabelle 1 dargestellt. Man sieht, daß in allen drei Einteilungen DIAL mit der Diagnose laut Krankengeschichte nahezu so gut wie FD2 und sogar besser als FD1 übereinstimmt, d.h. der Computer leistet dasselbe wie ein Fragebogendiagnostiker, dem die gleiche Information zur Verfügung steht. Es ist anzunehmen, daß sich die diagnostische Übereinstimmung erhöhen würde, wenn die CAPS vom jeweilig

Tabelle 1: Übereinstimmung der Erstdiagnosen

	KG-FD1	KG-FD2	KG-DIAL	FD1-DIAL	FD2-DIAL	FD1-FD2	TOTAL
Einteilung 1	73%	78%	74%	67%	70%	76%	73%
Einteilung 2	61%	69%	65%	61%	66%	70%	65%
Einteilung 3	54%	62%	54%	49%	53%	64%	56%

KG = Diagnose laut Krankengeschichte

DIAL = Computerdiagnose nach dem DIAL-Programm

FD1 und FD2 = Diagnosen der beiden "Fragebogendiagnostiker"

explorierenden Arzt selbst ausgefüllt worden wäre, da seine diagnostische Präferenz das Symptomprofil "überzeichnen" dürfte. Daß die beiden Fragebogendiagnostiker, die am selben Institut tätig sind, untereinander am besten übereinstimmen würden, war zu erwarten. Ihnen stand die gleiche Information zur Verfügung, wogegen sich die Krankengeschichtsdiagnose auch auf die Information aus dem persönlichen Kontakt mehrerer Ärzte mit dem Patienten und seinen Angehörigen, also andersartiges und umfangreicheres Material stützte. Der Computer dagegen arbeitete mit derselben beschränkten Information wie die Fragebogendiagnostiker, aber mit einem Programm, das von Fachleuten mit einer anderen Ausbildung auf einem anderen Kontinent entwickelt wurde.

Der Anteil übereinstimmend gewählter diagnostischer Kategorien wird mit zunehmender Differenzierung immer kleiner und ist bei Unterscheidung von effektiv 14 psychopathologischen Kategorien (Einteilung 3) um 15% bis 20% niedriger als bei Unterscheidung von nur vier Krankheitskategorien (Einteilung 1). Dies ist ebenfalls zu erwarten und wird grundsätzlich von allen Reliabilitätsprüfungen psychiatrischer Diagnosen berichtet (vgl. Zigler and Philipps³⁴). Die Übereinstimmung mit den Computerdiagnosen liegt dabei in Einteilung 3 um 2% bis 7% unter dem Gesamtdurchschnitt.

In der Annahme, daß bestimmte Diagnosen leicht und einstimmig gestellt werden, andere dagegen mehr Schwierigkeiten bereiten und abweichend gewählt werden, untersuchte ich als nächstes die Übereinstimmung innerhalb der diagnostischen Kategorien. Dabei wurden wieder nur die Hauptdiagnosen berücksichtigt (Tab. 2, 3 und 4).

Relativ am besten ist die Gesamtübereinstimmung bei Schizophrenien (durchschnittlich 51,0%); für Süchte (60,0%) variiert die paarweise Übereinstimmung beträchtlich (zwischen 38,0% und 88,0%), dasselbe gilt für Neurosen (17,0% bis 70,0%; durchschnittlich 44,0%). Die hohe Übereinstimmung bei sexuellen Deviationen (80,0%) und bei Manie (50,0%) sagt wegen des engen Zusammenhangs zwischen Leitsymptom und Diagnose und wegen der geringen Anzahl der Fälle nicht viel aus. Letzteres gilt auch für die

Tabelle 2: Übereinstimmung innerhalb der diagnostischen Kategorien, Einteilung 1

	KG - FD1	KG - FD2	KG - DIAL	FD1 - DIAL	FD2 - DIAL	FD1 - FD2	TOTAL
Psychosen	41 (29) 59%	41 (35) 70%	41 (35) 67%	37 (31) 60%	44 (36) 67%	37 (31) 62%	307 (197) 64%
Depression	18 (13) 43%	18 (14) 56%	18 (9) 45%	25 (8) 29%	21 (9) 39%	25 (16) 53%	156 (69) 44%
Neurosen etc.	39 (30) 67%	39 (27) 60%	39 (29) 62%	36 (27) 59%	33 (24) 52%	36 (27) 64%	271 (164) 61%
unklar	-	-	0 (0) 0%	0 (0) 0%	0 (0) 0%	-	12 (0) 0%

Zu jedem Diagnostikerpaa ist angegeben, wie oft jeder von ihnen die betreffende diagnostische Gruppe wählte; dazwischen in Klammern die Anzahl der Fälle, in denen diese Kategorie übereinstimmend von beiden gewählt wurde. Darunter Übereinstimmung in Prozent, bezogen auf die Häufigkeit, mit der die Kategorie von mindestens einem der beiden Diagnostiker gewählt wurde.

Tabelle 3: Übereinstimmung innerhalb der diagnostischen Kategorien, Einteilung 2

	KG - FD1	KG - FD2	KG - DIAL	FD1 - DIAL	FD2 - DIAL	FD1 - FD2	TOTAL
org. Psych.	11 (5) 14 25%	11 (5) 11 29%	11 (3) 8 19%	14 (6) 8 38%	11 (5) 8 36%	14 (9) 11 56%	99 (33) 33%
endog. Psych.	30 (16) 23 51%	30 (25) 33 66%	30 (25) 38 58%	23 (21) 38 53%	33 (27) 38 61%	23 (19) 33 51%	237 (135) 57%
Depression	18 (13) 25 43%	18 (14) 21 56%	18 (9) 11 45%	25 (8) 11 29%	21 (9) 11 39%	25 (16) 21 53%	156 (69) 44%
Neurosen etc.	32 (17) 20 49%	32 (18) 22 50%	32 (20) 27 51%	20 (15) 27 47%	22 (16) 27 48%	20 (15) 22 56%	202 (101) 50%
Süchte	7 (7) 16 44%	7 (6) 11 50%	7 (7) 10 70%	16 (10) 10 63%	11 (8) 10 62%	16 (10) 11 59%	84 (48) 57%
unklar	-	-	0 (0) 4 0%	0 (0) 4 0%	0 (0) 4 0%	-	12 (0) 0%

Legende vgl. Tabelle 2

TABLE 4. Distribution of Subjects by Age and Sex in the Study

	KG - FD1	KG - FD2	KG - DIAL	FD1 - DIAL	FD2 - DIAL	FD1 - FD2	TOTAL
Mania - Psych.	1 (1) 6 17%	1 (1) 5 20%	1 (0) 4 0%	6 (3) 4 43%	6 (3) 4 50%	6 (3) 5 57%	30 (12) 100%
Dep. Psych.	10 (4) 8 20%	10 (6) 6 28%	10 (3) 4 27%	8 (2) 4 20%	6 (2) 4 25%	8 (5) 6 37%	68 (12) 100%
Schizophrenia	26 (15) 19 30%	26 (20) 28 30%	26 (23) 35 61%	19 (18) 35 50%	28 (20) 35 41%	19 (12) 23 47%	172 (110) 100%
Psychose	3 (0) 1 0%	3 (0) 3 0%	3 (0) 2 0%	1 (0) 2 0%	3 (0) 2 0%	1 (0) 3 0%	26 (0) 0%
Manie	1 (1) 2 50%	1 (1) 2 50%	1 (0) 1 0%	2 (1) 1 50%	2 (1) 1 50%	2 (1) 2 100%	12 (0) 50%
Depression	18 (13) 25 43%	18 (14) 21 56%	18 (9) 11 45%	25 (8) 11 29%	21 (9) 11 39%	25 (16) 21 53%	150 (100) 44%
Neurosen	16 (12) 15 63%	16 (13) 14 76%	16 (7) 12 33%	15 (4) 12 17%	14 (5) 12 24%	15 (11) 14 61%	119 (80) 44%

(Fortsetzung von Tabelle 4: Übereinstimmung innerhalb der diagnostischen Kategorien, Einteilung 3)

	KG - FD ₁	KG - FD ₂	KG - DIAL	FD ₁ - DIAL	FD ₁ - FD ₂	TOTAL
Psychopathien	6 (0) 1 0%	6 (1) 1 17%	6 (1) 4 11%	1 (0) 4 0%	1 (0) 1 0%	33 (3) 9%
sex. Abw.	2 (2) 2 100%	2 (2) 3 67%	2 (2) 2 100%	3 (2) 2 67%	2 (2) 3 67%	15 (12) 80%
Alkoholabusus	4 (3) 8 33%	4 (2) 4 33%	4 (3) 4 60%	8 (4) 4 50%	8 (3) 4 33%	42 (18) 43%
sonst. Süchte	3 (3) 8 38%	3 (3) 7 43%	3 (3) 6 50%	8 (6) 6 75%	8 (7) 7 88%	45 (27) 60%
psychosomat. Krankheiten	3 (0) 0 0%	3 (0) 0 0%	3 (1) 6 13%	0 (0) 6 0%	0 (0) 0 -	26 (1) 4%
psych. Störung b. org. Krankh.	5 (0) 2 0%	5 (1) 4 13%	5 (1) 3 14%	2 (0) 3 0%	2 (0) 4 0%	39 (3) 8%
unklar	-	-	0 (0) 4 0%	0 (0) 4 0%	-	12 (0) 0%

Legende vgl. Tabelle 2

Nichtübereinstimmung bei den schizoaffectiven Psychosen (0%).

In der Kategorie "Depression" werden die Diagnosen psychotische und neurotische Depression mit reaktiver depressiver Psychose zusammengefaßt (Übereinstimmung 29,6 bis 56%; durchschnittlich 44,6); die Unterscheidung dieser Untergruppen gelingt nicht befriedigend. Viermal diagnostizierte der Computer eine Angstneurose in Fällen, wo Krankengeschichte und Fragebogendiagnostiker übereinstimmend eine Depression diagnostizierten (Bei der Faktorenanalyse der CAPPS bilden Angst und Depression einen gemeinsamen Faktor; vgl. S. 13). Eine Untersuchung von Sandifer²¹⁾, in der zehn Teilnehmer einer Fallkonferenz jeweils anschließend den betreffenden Patienten einer von 12 diagnostischen Kategorien zuzuordnen hatten, ergab bei affektiven Störungen ebenfalls sehr niedrige Übereinstimmung (Involutionspsychose 26%, manisch-depressive Psychose 36%).

Bei psychosomatischen Störungen und Psychopathien herrscht praktisch keine Übereinstimmung (4% bzw. 9%), wofür bereits Gründe angegeben wurden (vgl. S. 21). Die Nichtübereinstimmung für nicht-psychotische psychische Störungen bei körperlichen Erkrankungen korrespondiert mit der geringen Übereinstimmung bei organischen Psychosen (durchschnittlich 24,6): Häufig wird derselbe Patient z.B. von FD2 als psychisch gestörter, aber nicht-psychotischer "Organiker" und von FD1 als psychotischer "Organiker" eingestuft.

Schließlich wurden auch die Zweitdiagnosen verwendet, und zwar auf folgende Weise: Wenn beide Diagnostiker, etwa FD1 und FD2, nur je eine Diagnose gestellt hatten, konnte nur ein Diagnosenpaar verglichen werden. Wenn beide eine Erst- und eine Zweitdiagnose angegeben haben, lassen sich zwei Paare bilden. In Fällen jedoch, wo FD1 zwei und FD2 nur eine Diagnose gestellt hat, ging ich ebenfalls von zwei Paaren aus, denn die einzige Diagnose von FD2 kann mit jeder der beiden Diagnosen von FD1 verglichen werden. Außerdem ist eine nicht gestellte Zweitdiagnose in Fällen, wo der andere Diagnostiker eine Zusatzdiagnose für zutreffend erachtet hat, auch als eine Art "nicht übereinstimmender" Diagnose anzusehen. Bei

der Auszählung der übereinstimmenden Diagnosenpaare wurde jede einzelne Diagnose nicht öfter als einmal berücksichtigt. Durch diese Methode fielen die Werte niedriger aus als beim Vergleich der Erstdiagnosen (siehe Tab. 5).

Die Abhängigkeit der diagnostischen Übereinstimmung vom Sicherheitsgrad, mit dem die jeweilige Diagnose gestellt wurde, ist hier nicht berücksichtigt. Deck et al.⁴⁾ berichten, daß die Übereinstimmung zwischen Klinikern, die unabhängig voneinander dieselben Fälle explorierten und diagnostizierten, von 81% bei "Sicherheit" auf unter 50% bei "Wahrscheinlichkeit" und "Unsicherheit" der Diagnose absank. Sandifer²¹⁾ beschreibt den Effekt als weniger gravierend (Absinken von 65% auf 45%).

4. Zusammenfassung

DIAL, eine modifizierte Version des Computerprogramms DIAGNO II von Spitzer und Endicott, und die "Current and Past Psychopathology Scales" (CAPPS), ein von denselben Autoren entwickelter Fragebogen für die psychiatrische Exploration, wurden auf ihre diagnostische Verwendbarkeit in der deutschen Psychiatrie geprüft. Der Grad an Übereinstimmung zwischen der Krankenblattdiagnose von 98 Klinikpatienten und der Diagnose, die auf der Grundlage der CAPPS von zwei Klinikern und vom Computer jeweils unabhängig voneinander gestellt wurde, zeigt, daß das Verfahren auch hierzulande erfolgreich angewendet werden kann. Durch Verbesserung der Eingabeskalen und weitere Modifikationen des Entscheidungsprozesses könnte die Übereinstimmung zweifellos noch erhöht und ein für Ausbildungszwecke ausreichender Grad an diagnostischer Validität erreicht werden. Für die psychiatrische Ausbildung erscheint der bei DIAL mögliche Dialog zwischen Arzt und Computer (man-machine-dialogue) besonders geeignet³²⁾.

Tabelle 5: Übereinstimmung von Haupt- und Zusatzdiagnosen

	KG-FD1	KG-FD2	KG-DIAL	FD1-DIAL	FD2-DIAL	FD1-FD2
Einteilung 1	140 (81) 58%	169 (106) 63%	171 (96) 56%	136 (72) 53%	179 (100) 56%	145 (83) 57%
Einteilung 2	140 (69) 49%	169 (100) 59%	171 (88) 51%	136 (65) 48%	179 (89) 50%	145 (77) 53%
Einteilung 3	140 (61) 44%	169 (82) 48%	171 (68) 40%	136 (50) 37%	179 (66) 37%	145 (69) 48%

Angegeben ist jeweils die Anzahl möglicher Diagnosenpaare, dahinter in Klammern die Anzahl übereinstimmender Diagnosenpaare. Darunter Übereinstimmung in Prozent.

Literatur

1. American Psychiatric Association: Diagnostic and statistical manual: mental disorders. Washington: American Psychiatric Association 1952.
2. Ash, P.: The reliability of psychiatric diagnosis. J. Abnorm. Soc. Psychol. 44, 272-277 (1949).
3. Beck, A.T.: Reliability of psychiatric diagnosis: 1. A critique of systematic studies. Amer. J. Psychiat. 119, 210-216 (1962).
4. Beck, A.T., Ward, C.H., Mendelson, M., Mock, J.E., Erbaugh, J.K.: Reliability of psychiatric diagnosis: A study of consistency of clinical judgments and ratings. Amer. J. Psychiat. 119, 351-357 (1962).
5. Cassano, G.B.: Medical Tribune Bericht, Medical Tribune Nr.15: 3 (1972) und persönl. Mitteilung.
6. Cooper, J.E., Kendell, R.E., Gurland, B.J., Scharpe, L., Copeland, J.R.M., Simon, R.: Psychiatric diagnosis in New York and London. London-New York-Toronto: Oxford University Press 1972.
7. Degkwitz, R., Helmchen, H., Mombour, W. (Hrg.): Diagnoseschlüssel und Glossar psychiatrischer Krankheiten. Deutsche Übersetzung der internationalen Klassifikation der WHO (ICD), 8. Revision, und des internationalen Glossars. 3. Aufl. Berlin-Heidelberg-New York: Springer 1973.
8. Essen-Möller, E., Wohlfahrt, S.: Suggestions for the amendment of the official swedish classification of mental disorders. Acta psychiat. scand., suppl. 47, 551-555 (1947).
9. Gauron, E.F., Dickinson, J.K.: Diagnostic decision making in psychiatry: I. Information usage. Arch. Gen. Psychiat. 14, 225-232 (1966); II. Diagnostic styles. Arch. Gen. Psychiat. 14, 233-237 (1966).
10. Gerster, F., Dirlich, G.: DIAL, ein Programm zur Entscheidungshilfe in der Psychiatrie für das IBM-System 1130. München: im Druck.

11. Gross, R.: Medizinische Diagnostik - Grundlagen und Praxis. Berlin-Heidelberg-New York: Springer 1969.
12. Krauß, B.: Empirische Untersuchung zur Verlässlichkeit psychiatrischer Diagnosen. Med. Diss., Univ. München (in Vorbereitung).
13. Kreitman, N., Sainsbury, P., Morrissey, J., Towers, J., Scrivener, J.: The reliability of psychiatric assessment: An analysis. J. Ment. Sci. 107, 887-908 (1961).
14. Lange, H.J.: Algorithmische Diagnostik. EDV, BMT, NM 2, 577-580 (1971).
15. Lorr, M., Klett, C.J., McNair, D.M., Lasky, J.J.: Inpatient multidimensional psychiatric scale (IMPS). Palo Alto/Calif.: Consulting Psychologists Press 1962.
16. Meyer, J.E.: Diagnostische Einteilungen und Diagnosenschemata in der Psychiatrie. In: H.W. Gruhle, R. Jung, W. Mayer-Gross u. M. Müller: Psychiatrie der Gegenwart, Bd. 3, S. 130-180. Berlin-Göttingen-Heidelberg: Springer 1961.
17. Mombour, W.: Klassifikation, Patientenstatistik, Register. In: K.P. Kisker, J.E. Meyer, M. Müller, E. Strömberg: Psychiatrie der Gegenwart. Bd. 3, 2. Aufl. Berlin-Heidelberg-New York: Springer (im Druck).
18. Mombour, W.: Verfahren zur Standardisierung des psychopathologischen Befundes. Psychiat. clin. 5, 73-120 und 137-157 (1972).
19. Nathan, P.E.: Cues, decisions, and diagnoses. New York-London: Academic Press.
20. Overall, J.E., Gorham, D.R.: A pattern probability model for the classification of psychiatric patients. Behav. Sci. 8, 108-116 (1963).
21. Sandifer, H.G., Pettus, C., Quade, D.: A study of psychiatric diagnosis. J. nerv. ment. Dis. 139, 350-356 (1964).
22. Scharfetter, O.: Das AMP-System. 2. Aufl. Berlin-Heidelberg-New York: Springer 1972.
23. Schmidt, H.O., Fonda, C.P.: The reliability of psychiatric diagnosis: A new look. J. Abnorm. Soc. Psychol. 52, 262-267 (1956).

24. Shepherd, M., Brooke, E.M., Cooper, J.E., Lin, T.: An experimental approach to psychiatric diagnosis. *Acta psychiat. scand.*, suppl. 201 ad vol. 44 (1968).
25. Smith, W.G.: A model for psychiatric diagnosis. *Arch. gen. Psychiat.* 14, 521-529 (1966).
26. Spitzer, R.L., Cohen, J., Fleiss, J.L., Endicott, J.: Quantification of agreement in psychiatric diagnosis: A new approach. *Arch. gen. Psychiat.* 17, 83-87 (1967).
27. Spitzer, R.L., Endicott, J.: DIAGNO: A computer program for psychiatric diagnosis utilizing the differential diagnostic procedure. *Arch. gen. Psychiat.* 18, 746-756 (1968).
28. Spitzer, R.L., Endicott, J.: DIAGNO II: Further developments in a computer program for psychiatric diagnosis. *Amer. J. Psychiat.* 125, suppl. 12-21 (1969).
29. Ward, C.H., Beck, A.T., Mendelson, M., Mock, J.E., Erbaugh, J.K.: The psychiatric nomenclature: Reasons for diagnostic disagreement. *Arch. gen. Psychiat.* 7, 198-205 (1962).
30. Wing, J.K., Birley, J.L.T., Cooper, J.E., Graham, P., Isaacs, A.D.: Reliability of a procedure for measuring and classifying "present psychiatric state". *Brit. J. Psychiat.* 113, 499-515 (1967).
31. Wing, J.K.: A standard form of psychiatric present state examination. In: *Psychiatric Epidemiology*, ed. by Hare, E.H., and Wing, J.K. London-New York-Toronto: Oxford University Press 1970.
32. Zerssen, D.v.: Vorschläge für die klinische Ausbildung in psychiatrischer Diagnostik. 87. Wanderversammlung Südwestdeutscher Psychiater, Baden-Baden 1971. Ref.: *Zbl. ges. Neurol. u. Psychiat.* 204, Heft 2, Ref. 546 (1972).
33. Zerssen, D.v.: Diagnose. Nosologie. In C. Müller: *Lexikon der Psychiatrie*. Berlin-Heidelberg-New York: Springer 1973.
34. Zigler, E., Philipps, L.: Psychiatric diagnosis: A critique. *J. abnorm. soc. Psychol.* 63, 607-618 (1961).
35. Zubin, J.: Classification of the behavior disorders. *Ann. Rev. Psychol.* 18, 373-406 (1967).

Lebenslauf

Ich wurde am 19. Mai 1949 in Prien am Chiemsee geboren, wo ich ab 1955 die Grundschule und ab 1959 das Ludwig-Thoma-Gymnasium besuchte. 1968 legte ich die Reifeprüfung ab und begann im Wintersemester 1968/69 mit dem Studium der Humanmedizin an der Ludwig-Maximilians-Universität in München.

Im Sommer 1969 absolvierte ich die naturwissenschaftliche Vorprüfung und im Winter 1970/71 die ärztliche Vorprüfung. Von Oktober 1972 bis Juli 1973 studierte ich mit einem Auslandsstipendium der Studienstiftung des deutschen Volkes an der Medical School der University of Birmingham in England.

Nach dem Sommersemester 1974 möchte ich in München das Staatsexamen ablegen.

